Available online at www.sciencedirect.com

SCIENCE DIRECT®

Virology 321 (2004) 29–35

VIROLOGY

www.elsevier.com/locate/yviro

Crimean–Congo hemorrhagic fever virus genome L RNA segment and encoded protein

Jessica E. Honig, Jane C. Osborne, and Stuart T. Nichol*

Special Pathogens Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

Received 3 September 2003; returned to author for revision 16 September 2003; accepted 18 September 2003

Abstract

Sequence analysis of the L RNA genome segment and predicted encoded L polymerase protein of Crimean–Congo hemorrhagic fever (CCHF) virus (genus *Nairovirus*, family *Bunyaviridae*) demonstrates that they are approximately twice the size of those found in viruses of other bunyavirus genera. The CCHF virus L segment and encoded protein (12164 nucleotides and 3944 amino acids, respectively) are similar in size and sequence to those of the nairovirus Dugbe virus (12255/62% and 4036/62% nucleotide and amino acid length/identity, respectively). The identification of an ovarian tumor (OTU)-like protease motif in the L protein amino termini of the nairoviruses Dugbe, CCHF, and Nairobi sheep disease (NSD) indicates these proteins are members of the recently described OTU-like protease family and suggests that these large proteins may be polyproteins that are autoproteolytically cleaved or involved in deubiquitination.
Published by Elsevier Inc.

Keywords: Nairovirus; Virus; Crimean; Congo; Hemorrhagic; OTU; Polymerase; Dugbe; L; Deubiquitination

Introduction

The Crimean–Congo hemorrhagic fever (CCHF) virus, genus *Nairovirus*, family *Bunyaviridae*, is the tick-borne causative agent of outbreaks of severe hemorrhagic fever with high mortality in regions of Asia, Europe, and Africa (Nichol, 2001; Swanepoel, 1994). As a result, CCHF virus has been classified as a category A select biological agent and is designated as a biological safety level four pathogen. The nairovirus genome is composed of a negative-stranded RNA that exists as three segments designated S, M, and L (Marriott and Nuttall, 1996; Schmaljohn and Hooper, 2001). The only nairovirus genome to have been fully sequenced thus far is that of Dugbe (DUG) virus. The DUG virus L segment is 12255 nucleotides long, containing a single open reading frame encoding a protein of 459 kDa that is predicted to be the viral polymerase, L (Marriott and Nuttall, 1996). Hence, DUG virus L is almost twice the size of that of many other members of the family *Bunyaviridae* (Fig. 1). While S and M segment sequence data exist for CCHF virus, the absence of complete L sequence data has been a major hurdle in developing techniques to further study the virus, including attempts to generate an infectious clone (Flick et al., 2003).

Results

RT-PCR and sequencing strategy

To design primers that amplify regions of the CCHF L segment we used the DUG virus L polymerase amino acid sequence as the query sequence in a position-specific iterated (PSI)-BLAST search of all sequences available in the nr (all nonredundant GenBank CDS translations + RefSeq Proteins + PDB + SwissProt + PIR + PRF) database available at the NCBI website (http://www.ncbi.nlm.nih.gov/). After several iterations, multiple conserved short amino acid motifs were identified among the polymerase proteins of single-stranded segmented RNA viruses including members of the *Bunyaviridae* and *Arenaviridae* families, Tenuiviruses, and PB1 proteins of the influenza viruses (data not shown). These included the previously identified conserved polymerase core motifs described by others

* Corresponding author. Special Pathogens Branch, Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Mailstop G14, 1600, Clifton Road, N.E., Atlanta, GA 30333. Fax: +1-404-639-1118.

E-mail address: snichol@cdc.gov (S.T. Nichol).

0042-6822/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.virol.2003.09.042

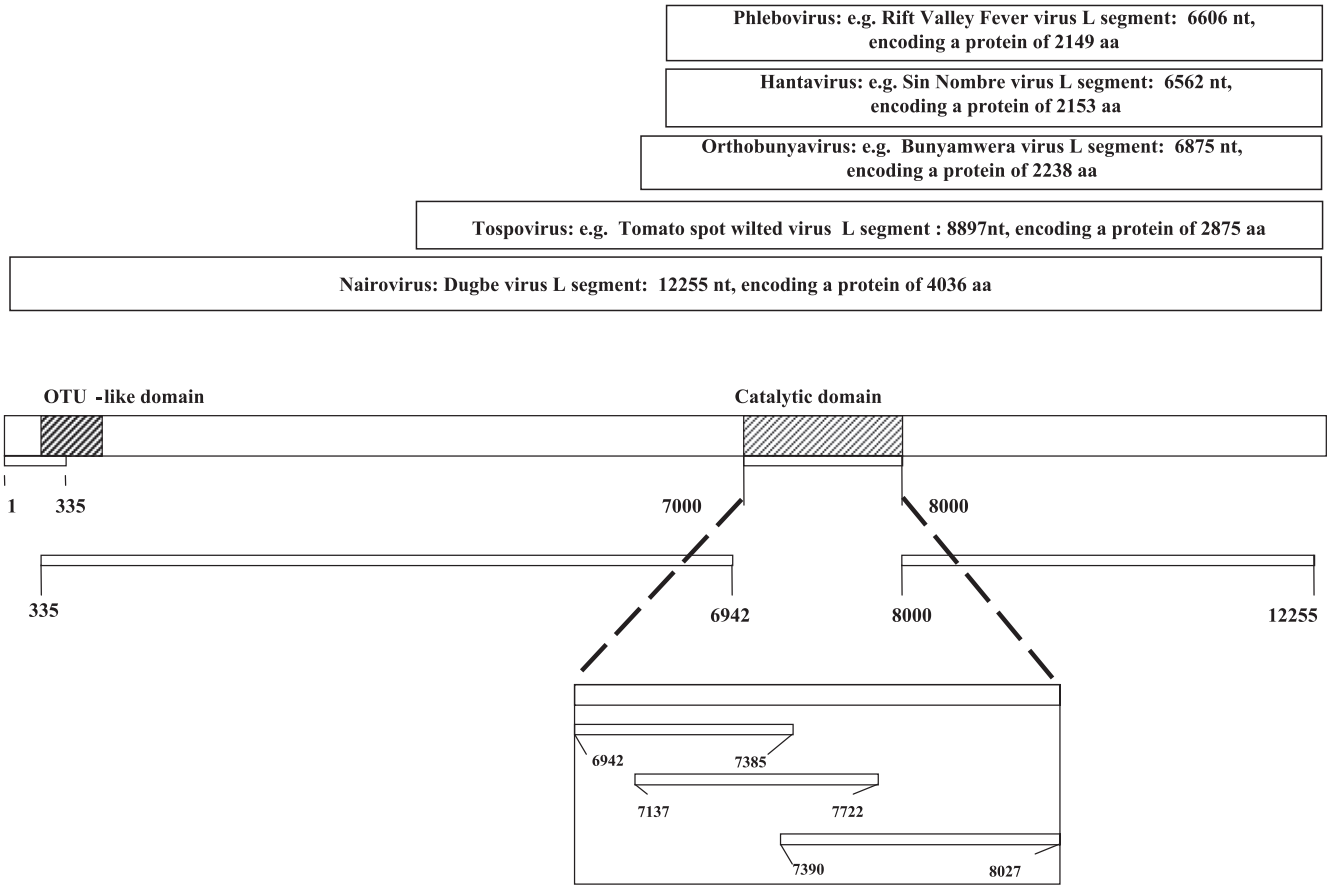

Fig. 1. Dugbe virus L RNA segment and encoded polymerase sequence features utilized in the design of the CCHF virus L sequencing strategy. Relative sizes of the L RNA segment-encoded proteins of viruses of each of the five genera of the family *Bunyaviridae* are shown in the upper panel. Length differences locate predominantly at the amino terminus of the proteins. The location of conserved motifs in DUG L utilized in initial primer design and CCHF L amplification strategy is shown in the lower panel. The positions of primers are indicated.

(Aquino et al., 2003; Muller et al., 1994). In addition, we discovered an ovarian tumor (OTU)-like cysteine protease motif present within the sequence close to the amino terminus of the Dugbe virus L protein (Fig. 2). Using these relatively conserved motifs as a guide, sets of oligonucleotide primers were designed (Fig. 1, and see Materials and methods) to the putative polymerase catalytic region (primer pairs 6942 and 7385, 7137 and 7722, 7390 and 8027) and to the newly predicted OTU-like protease domain (primer 335) (Fig. 2). A degenerate primer was designed to the termini of the CCHF L segment by comparison of the conserved positions of the 5′ and 3′ S, M, and L segment terminal sequences of DUG virus and the S and M segment terminal sequences of CCHF virus (primer 5/1/20).

Three sets of RT-PCR reactions were performed (Fig. 1). The first involved amplification of the section of CCHF L between the 5′ positive-sense terminus and the region corresponding to the amino terminus of the L protein (primers 5/1/20 and 335). Sequencing of this product allowed identification of the start codon of the L protein and construction of a nondegenerate primer to this sequence. The second set of RT-PCR reactions amplified a region of approximately 1 kb encompassing the putative conserved polymerase catalytic domain. Sequencing of this product led to the design of nondegenerate primers to both extremes of the amplified region. In the third reaction, the region corresponding to the sequence between the carboxy terminus of the conserved polymerase catalytic domain and the 3′ positive-sense terminus was amplified (primers 8000 and 5/1/20). Finally, the region between the sequences corresponding to the start of the gene and the amino-terminal end of the catalytic domain was reverse-transcribed and amplified. Sequencing was then performed in both directions on the RT-PCR products, and novel sequencing primers designed as new sequence became available. A consensus genome sequence was developed based on multiple overlapping PCR products and sequencing reads.

Segment and sequence features

The CCHF virus L RNA segment is 12164 nucleotides long, comprising a 5′ noncoding region of 76 nucleotides, L protein open reading frame of 11835 nucleotides, and a 3′ noncoding region of 253 nucleotides long (Genbank accession number AY389508). The L ORF would encode a protein product of 3944 amino acids and 448 kDa, making

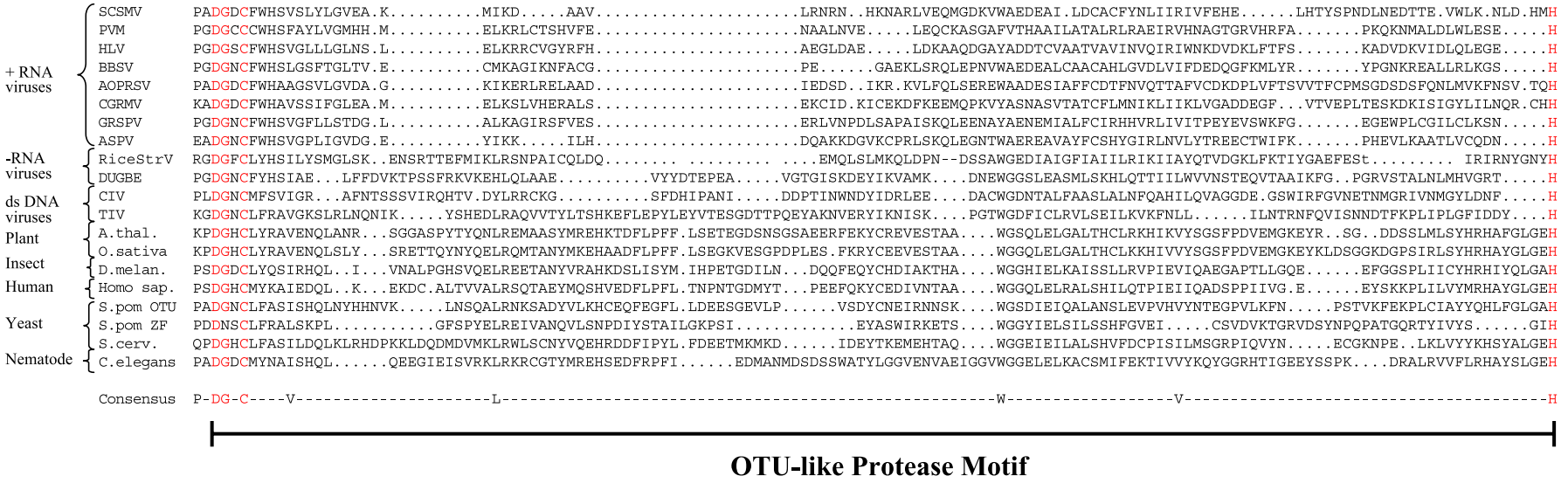

Fig. 2. Dugbe virus L RNA segment-encoded protein belongs to a family of OTU-like cysteine proteases. Representatives of proteins containing an OTU-like protease motif. These include the RNA polymerase containing polyprotein encoded by positive strand RNA viruses such as Carlaviruses: SCSMV, Sugar cane striate mosaic virus (Genbank aad10481); PVM, Potato virus M (PN0093); HLV, Hop Latent virus (BAB13712); BBSV, Blueberry scorch virus (AAA68984) and the Foveaviruses: AOPRSV, African Oil Palm Ring Spot virus (aal68924); CGRMV, Cherry Green Ring Mottle virus (AAC35433); GRSPV, Grapevine Rupestris stem pitting virus (AAC62910); ASPV, Apple stem pitting virus (BAB15950). Also included is a representative of the RNA polymerase encoded by the negative strand RNA viruses of the Tenuivirus group, for example, RiceStrV, Rice stripe virus (gi563778) and Dugbe virus, genus *Nairovirus*, family *Bunyaviridae*, DUGBE (gi3914897). Double stranded DNA viruses Iridoviruses are included; for example, CIV, Chilo iridescent virus (AAK82093) and TIV, *Tipula iridescent* virus (AAA47919). Such proteins are also numerous among plants, for example, A.thal, *Arabidopsis thaliana* putative protein (CAB87739) and O.sativa, *Oryza sativa* putative protein (CAB51835), and invertebrates, for example, *Drosophila*, D.melanogaster CG7857 OTU-like (AAF49636), and yeast, for example, S.cerevisiae (AAB65066), S.pombe OTU-like cysteine protease (CAB52567); S.pombe hypoth. Zinc finger protein (CAB11271). This motif can also be seen in humans, for example, Homo sap. CGI-77 (AAD34073) and nematodes, for example, C.elegans (CAA91033).

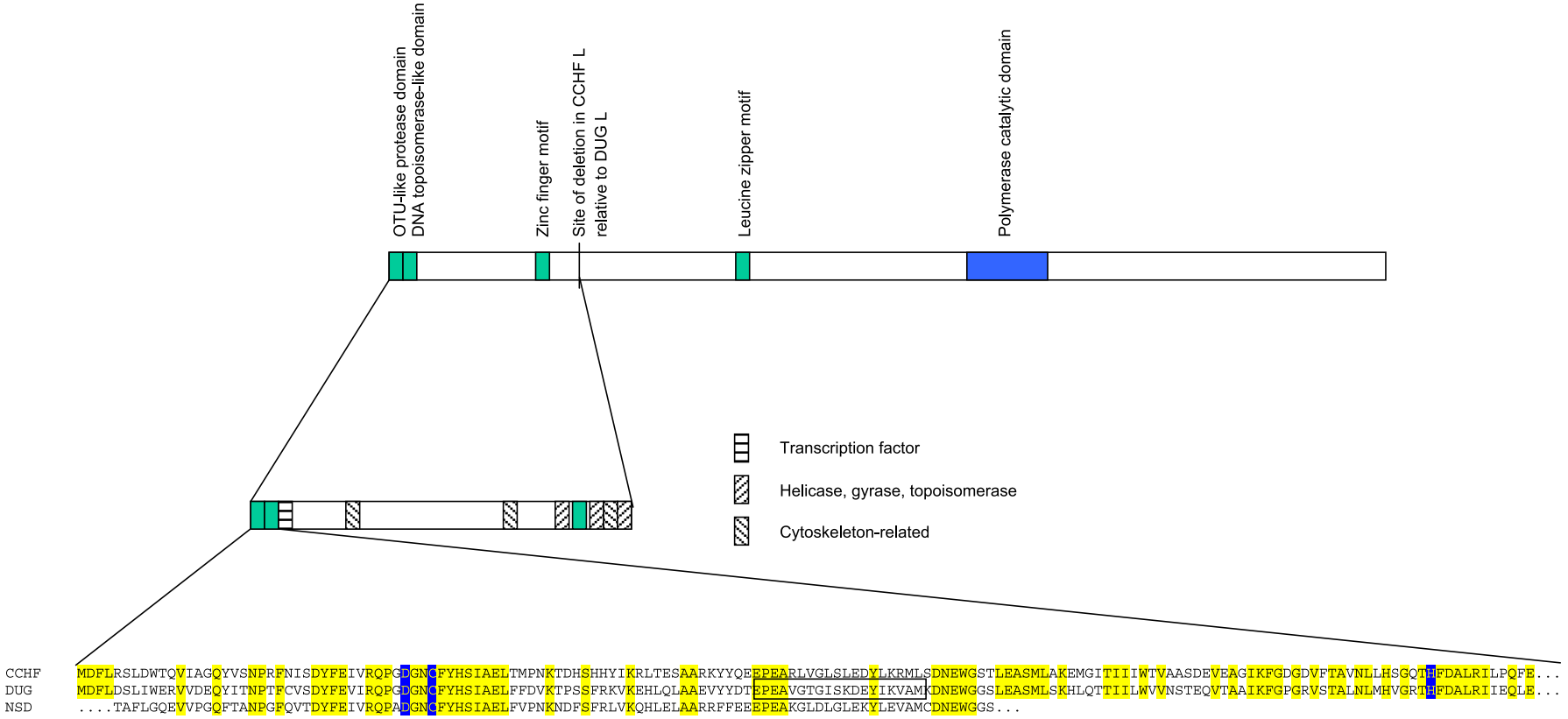

Fig. 3. Sequence motifs present in CCHF L RNA segment-encoded proteins. The upper panel depicts the approximate location of various sequence motifs predicted within the entire CCHF L encoded protein (shown from the N terminus to the C terminus) along with an enlargement of the amino terminal region displaying amino acid identity with proteins involved with the cytoskeleton and nucleic acid interactions. The site of a 92 amino acid deletion in CCHF L relative to DUG L is also shown. The lower panel shows the alignment of the amino terminus of the L encoded protein of CCHF, Dugbe, and NSD (partial sequence) viruses. Conserved positions are shown in yellow. The highly conserved D, C, and H residues believed to constitute the OTU-like protease core amino acid motif are shown in blue. The boxed region within the DUG virus sequence (bottom panel) indicates the position of the predicted topoisomerase I active site motif. Highly similar sequences can be seen in both CCHF and NSD viruses.

it comparable with DUG virus L segment which is 12255 nucleotides long, encoding a protein of 4036 amino acids and 459 kDa. CCHF L has a 36 nt insertion in the 5′ noncoding region relative to that of Dugbe virus. Also of note is a 92 amino acid deletion in CCHF L protein relative to DUG L in the region of CCHF L amino acids 756–799. The overall degree of sequence identity with DUG L segment was calculated to be 62% both at the nucleotide and amino acid levels, with the most highly conserved area being that encoding the region from approximately amino acids 2050 to 2700, which corresponds to the conserved core catalytic domains of RNA-dependent RNA polymerases that have also been predicted to be present in DUG virus L protein (Aquino et al., 2003; Marriott and Nuttall, 1996; Muller et al., 1994).

Further analysis of the sequences of CCHF L and DUG L using the PredictProtein server (http://maple.bioc.columbia.edu/pp/) indicated the presence of a C2H2-type zinc finger domain (amino acid positions: DUG 608–631; CCHF 609–632) and a leucine zipper motif (DUG 1472–1493; CCHF 1386–1407) (Fig. 3). A sequence in the CCHF L was also highly similar to a eukaryotic DNA topoisomerase I active site motif predicted at the amino terminus of DUG L (aa 76–94).

During the initial primer design phase of the project, other potential functional domains were identified during PSI-BLAST search of all sequences available in Genbank. Of particular note is the presence of an OTU-like protease motif that is predicted in the extreme amino terminus of DUG L and CCHF L protein (Figs. 2 and 3). Elements of this motif were also present in a partial Nairobi Sheep Disease (NSD) virus L sequence fragment generated (Fig. 3). A BLAST search for short regions of high homology to other proteins yielded several results indicating homology between the deduced CCHF L amino-terminal 750 residues and proteins involved in nucleic acid interactions (helicase, gyrase, and topoisomerase proteins, and transcription factors) and cytoskeleton-associated proteins (Fig. 3).

Discussion

CCHF virus is an important human pathogen. Research on the virus has been hindered by the absence of sequence data on the virus genome RNA L segment. This report constitutes the completion of sequencing of the CCHF virus genome and represents the final hemorrhagic fever select agent gene sequence to be ascertained. Considering the degree of sequence identity between the CCHF L and DUG L and the presence of the conserved polymerase motifs, CCHF L is predicted to encode the viral RNA-dependent RNA polymerase. Based on the BLAST search results and phylogenetic analysis (data not shown), the nairovirus RNA-dependent RNA polymerase catalytic core is clearly more closely related to that of the tenuiviruses and arenaviruses than to those of members of other genera of the family *Bunyaviridae*. The lack of monophyly of the polymerases of the family *Bunyaviridae* viruses shows that their L RNA segments lack a common ancestor and indicate a deeper shared ancestry with other negative-stranded segmented RNA viruses including the arenaviruses, tenuiviruses, and influenza PB1 proteins.

The amino termini (amino acids 34–152) of the L gene product of DUG, CCHF, and NSD viruses are predicted to harbor a conserved OTU-like protease motif, a predicted papain-like protease with a cysteine protease signature (Figs. 2 and 3). This finding is particularly important as it indicates that the nairovirus L segment encoded proteins are part of a larger family of OTU-like proteases. This protein family includes proteins encoded by segmented negative strand viruses belonging to the tenuivirus group (e.g., rice stripe virus polymerase amino acids 39–150), and polyproteins encoded by several positive strand RNA viruses (carlaviruses and foveaviruses) which are autoproteolytically cleaved to generate the viral RNA-dependent RNA polymerase and additional proteins such as viral helicases (Lawrence et al., 1995). This protein family has recently been annotated in the NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cddfind.cgi), and unbeknown to us at the time our work was performed, had been described earlier in a review article (Makarova et al., 2000). The polyproteins of carlaviruses undergo autoproteolytic cleavage to generate NTP/helicase and polymerase domains, attributable to a domain in which the OTU-like protease motif is predicted (Lawrence et al., 1995). This is interesting considering that BLAST sequence analysis indicated that the amino-terminal 750 amino acids of nairovirus L proteins display homology to helicases, gyrases, and topoisomerases. While the proteolytic function of the OTU-like protease motif has not been shown, it is possible that it functions in the nairovirus L protein by autoproteolytically cleaving the polyprotein to yield a polymerase and helicase. This hypothesis is supported by the fact that nairovirus L open reading frames are almost twice the size of those of the other members of the family *Bunyaviridae,* suggesting that they could encode other proteins in addition to the viral polymerase. Homology of the amino terminus of CCHF L protein with transcription factors and cytoskeleton-associated proteins suggests that the polyprotein may have other functions.

There have recently been reports linking the OTU-like protease superfamily with deubiquitination activity (Balakirev et al., 2003; Borodovsky et al., 2002). Some OTU-like protease-containing proteins that have been implicated in the ubiquitination pathway have been found to contain zinc-finger-like motifs (Makarova et al., 2000). This is particularly interesting considering a zinc-finger motif is predicted at amino acids 608–632 of both CCHF and DUG L proteins. While CCHH L amino acids 1545–1611 show identity to ubiquitin-conjugating enzyme E2, neither CCHF nor DUG L is predicted to possess the putative ubiquitin-interacting motif or ubiquitin-associated domain

identified in some deubiquitinating proteins that contain OTU-like protease motifs (Balakirev et al., 2003; Hofman and Bucher, 1996; Hofman and Falquet, 2001). Further study is needed to determine whether nairovirus L proteins are directly involved in deubiquitination and what role this may play in the virus replication cycle. Deubiquitinating activity of a viral protease has recently been demonstrated for adenovirus L3 23K proteinase (Balakirev et al., 2002). In addition, regulation of herpes simplex virus type 1 gene expression has been shown to involve the interaction of a viral protein with HAUSP, a ubiquitin-specific protease (Everett et al., 1997).

Materials and methods

Virus strains and origins

CCHF virus strain IbAr10200 (originally isolated from *Hyalomma excavatum* ticks from Sokoto, Nigeria, in 1966) and NSD virus isolate RV082 (from Kenya) were used in this study. All work with CCHF virus was performed in the biosafety level four laboratory at the Centers for Disease Control and Prevention, Atlanta. NSD virus RNA was provided by Cindy Rossi (USAMRIID, Fort Detrick, MD).

RNA purification, RT-PCR, and sequencing

CCHF or NSD virus infected Vero E6 cells and supernatant were harvested, freeze–thawed, and total RNA was extracted from 0.2 ml medium with 1 ml of isolation reagent (Tripure; Roche). Two hundred microliters of chloroform was added, and RNA was purified from the aqueous phase by use of an RNaid kit (Q-Biogene). Initial RT-PCR reactions were performed using the degenerate primer pairs (shown 5′to 3′): ATGATTGCIAAYAGIAAYTTYAA (6942) + ACAGCARTGIATIGGICCCCAYTT (7385); CCCAARGCICARCTIGGIGG (7137) + TGAAGAAGTIGCATGRTGIATICCYT (7722); ATATCTGGIGAYAAYACIAARTGG (7390) + CATCATAAATTCICTITARAAYTC (8027); TCTCAAAGAIATCIITCCCC (5/1/20) + GAAGCYTCYAAIGAICCICCCCA (335); TCTCAAAGAIATCIITCCCC (5/1/20) + GTAGTTCAGATGATTACGC (8000).

Most reactions were performed using the Access Quick RT-PCR system in a Techgene thermal cycler (Techne) according to the manufacturer's protocol (Promega). Reverse transcription was performed at 42 °C for 60 min and reverse transcriptase inactivated at 94 °C for 2 min. PCR reactions were performed using 40 cycles of a temperature profile of 94 °C for 30 s, ramping (over the 40 cycles) from 42 to 50 °C for 60 s, and then 68 °C for 1 min per kb, followed by a final extension at 68 °C for 7 min. Generation of the 7 kb RT-PCR product was performed using Thermoscript reverse transcriptase (Invitrogen), followed by amplification with Platinum Taq DNA polymerase (Invitrogen) following the manufacturer's instructions. Reverse transcription was carried out using a temperature gradient of 45–65 °C for 2 h in a gradient thermal cycler (DNA Engine, MJ Opticon). Two microliters of the cDNA was then amplified without purification using 40 cycles of a temperature profile of 94 °C for 30 s, 53 °C for 60 s, and 68 °C for 8 min, followed by a final extension at 68 °C for 20 min, in a Techgene thermal cycler (Techne). Each reaction was electrophoresed on a 1% agarose gel in 1× Tris borate-EDTA buffer containing ethidium bromide. RT-PCR products were visualized by UV transillumination. The PCR products were then purified from the gel by excision of the gel fragment, freezing, and spinning through a spin filter (Qiagen). After phenol/chloroform extraction, the DNA was ethanol precipitated.

Sequencing reactions were performed in both directions using ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kits (Applied Biosystems) according to the manufacturer's instructions. Dye was removed using DyeEx spin columns (Qiagen), and samples were sequenced on an ABI 3100 automated DNA capillary sequencer (Applied Biosystems). Sequences were analyzed with Sequencher 4.0.5 software (Gene Codes). The cDNA nucleotide sequences were used to search the GenBank database using BLAST software available on the NCBI website (Altschul et al., 1997; http://www.ncbi.nlm.nih.gov/BLAST/). Virus sequences were aligned with those of DUG virus using the PILEUP program of the Wisconsin Package Version 10.2 (Genetics Computer Group, Inc.).

Acknowledgments

We thank Beate Marczinke, Pierre Rollin, and Tom Ksiazek for their advice and valuable contributions to this project. We also thank Cindy Rossi (USAMRIID, Fort Detrick, MD) for providing inactivated RNA from NSD virus.

References

Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Aquino, V.H., Moreli, M.L., Morales Figueiredo, L.T., 2003. Analysis of Oropouche virus L protein amino acid sequence showed the presence of an additional conserved region that could harbour an important role for the polymerase activity. Arch. Virol. 148, 19–28.

Balakirev, M.Y., Jaquinod, M., Haas, A.L., Chroboczek, J., 2002. Deubiquitinating function of adenovirus proteinase. J. Virol. 76, 6323–6331.

Balakirev, M.Y., Tcherniuk, S.O., Jaquinod, M., Chroboczek, J., 2003. Otubains: a new family of cysteine proteases in the ubiquitin pathway. EMBO Rep. 4, 517–522.

Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K.D., Ploegh, H.L., Kessler, B.M., 2002. Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. Chem. Biol. 9, 1149–1159.

Everett, R.D., Meredith, M., Orr, A., Cross, A., Kathoria, M., Parkinson, J.,

1997. A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regulatory protein. EMBO J. 16, 1519–1530.

Flick, R., Flick, K., Feldmann, H., Elgh, F., 2003. Reverse genetics for Crimean–Congo hemorrhagic fever virus. J. Virol. 77, 5997–6006.

Hofman, K., Bucher, P., 1996. The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. Trends Biochem. Sci. 21, 172–173.

Hofman, K., Falquet, L., 2001. A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. Trends Biochem. Sci. 26, 347–350.

Lawrence, D.M., Rozanov, M.N., Hillman, B.I., 1995. Autocatalytic processing of the 223-kDa protein of blueberry scorch carlavirus by a papain-like proteinase. Virology 207, 127–135.

Makarova, K.S., Aravind, L., Koonin, E.V., 2000. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. Trends Biochem. Sci. 25, 50–52.

Marriott, A.C., Nuttall, P.A., 1996. Large RNA segment of Dugbe virus encodes the putative RNA polymerase. J. Gen. Virol. 77, 1775–1780.

Muller, R., Poch, O., Delarue, M., Bishop, D.H., Bouloy, M., 1994. Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. J. Gen. Virol. 75, 1345–1352.

Nichol, S.T., 2001. Bunyaviruses. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, vol. 2. Lippincott Williams and Wilkins, Philadelphia, PA, pp. 1603–1634.

Schmaljohn, C.S., Hooper, J.W., 2001. *Bunyaviridae*. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, vol. 2. Lippincott Williams & Wilkins, Philadelphia, PA, pp. 1581–1602.

Swanepoel, R., 1994. Crimean–Congo haemorrhagic fever. In: Coetzer, J.A.W., Thomson, G.R., Tustin, R.C. (Eds.), Infectious Diseases of Livestock with special reference to Southern Africa. Oxford Univ. Press, Cape Town, South Africa, pp. 723–729.